

DATA ANALYSIS OF BIOLOGICAL TIME SERIES

M.D.GODFREY

Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey, USA

H.HOFFMAN, R.MADDEN, C.S.PITTENDRIGH

Princeton University, Princeton, New Jersey, USA

and

S.SKOPIK

University of Delaware, Newark, Delaware, USA

Abstract: This paper presents some of the results of analyses of body temperature data of two species of mammals. The animals were subjected to various environmental conditions with respect to variations in periodic light cycle with a period close to 24 hr. New statistical techniques are used to study the animals as a function of time and change in environment. These techniques confirm some previously observed behavior and indicate new aspects of the dynamics of adjustment to change in photoperiod.

1. Introduction

This paper reports very briefly on the analysis of a large body of data which were recorded at Princeton University during the last several years. The data consist of temperature and activity recordings, taken at either two- or five-min intervals, on several species of small mammals. The animals were isolated in cages which were subjected to a variety of controlled photoperiods.

Tables 1–4 show the total lengths of the data recordings and the various conditions which were imposed.

Due to the limitation on the length of this paper, and also due to the fact that a full report is currently in preparation, we will not give any discussion of the important questions of the data recording techniques, or the data preparation and computer preprocessing which were carried out. We will simply state here that it appears that the resulting data which were finally analyzed by the techniques discussed in this paper were of high quality.

In section 2 we will give a relatively extended discussion of the statistical techniques which we used in the study of these data. Some of these techniques have not

Exp. 200

Table 1

Animal type	Channel	10	20	30	40	50	60	70	80	90	days
Ref.	1	LD 12:12			DD						
P	2	LD 12:12			DD						
P	3	LD 12:12			DD						
M	4	LD 12:12			12:12 skeleton photoperiod (15 min light each 12 hr)						
P	5	LD 12:12			DD						
P	6	LD 12:12			DD						
M	7	LD 12:12			DD						
P	8	LD 12:12			DD						
P	9										
	10										
P	11	LD 12:12			DD						
P	12	LD 12:12			DD						
P	13	DD			DD						
P	14	DD									
	15										
M	16				DD						
Ref.	17				DD						
	18										

Animal type

P: *Perognathus formosus*M: *Mesocricetus auratus*P (bardi): *Peromyscus bardi*

Ref: reference telemeter (recording telemeter in empty cage)

Conditions

LD: light-dark cycle

LD 12.5:12: 12.5 hr of light then 12 hr darkness

DD: constant darkness

LL: constant light

T: 15 min light pulse each cycle

T₂₃: 15 min light each 23 hr.

Exp. 300

Table 2

Animal type	Channel	10	20	30	40	50	60	70	80	90	100 days
Ref.	1				DD	Ref T.					
P	2	LD 12:12	DD		LL						
P	3	LD 12:12	DD		LD 18:6						
M	4	LD 12:12			DD	X 1 hr light	DD				
P	5	LD 12:12	DD		12 hr light	DD					
	6										
M	7	LD 12:12			X 12 hr light	DD					
P	8	LD 12:12	X 1 hr light	DD						T ₂₂	
Ref.	9	LD 12:12			DD						
P	10	LD 12:12	X 1 hr light	DD							
P	11	LD 12:12	DD		LD 7:17						
	12										
M	13	LD 12:12			DD					T ₂₂	
M	14	LD 12:12	LD 14:10							DD	
M	15				DD						
M	16				DD						
	17										
M	18	LD 12:12			LL						

previously been applied in biology, and, in fact, have only been developed within the last few years.

In section 3, we will give a brief summary of the main results of our analyses. This summary is condensed and preliminary in nature. Again, a full discussion and further analysis are in preparation.

2. Methods

The purpose of this section is to describe the statistical techniques which we have used in the analysis of these time series data. Since much of the interest in the data concerns evidence of changes in the properties of the mechanism which generated the data, some novel statistical techniques were required. In particular, we have used techniques which do not require that the process remain invariant under time translations. With these techniques we have been able to study the transitions which occur when an animal's environment is suddenly changed. We have also, of course, applied the somewhat more conventional techniques of time series analysis to appropriately selected portions of our data. In general, we have used the techniques which do not depend on long time averages to investigate the appropriateness of forming such long averages. These averages are required, as will be discussed in greater detail below, in order to form precise estimates of any indication of periodic variation in the data.

We will first give a brief discussion of the computation of periodogram and spectrum estimates and their uses and limitations when applied to the study of nearly periodic phenomena. In the second part of this section we will discuss the computation and applications of frequency decomposition by use of complex-demodulates. The material in this later part will probably be less familiar to many readers. An effort will be made to give a fairly self-contained exposition of these data-analysis procedures.

2.1. Periodograms and spectra

In this section we will discuss the computation of periodograms and spectra from the standpoint of the statistical data analysis implications of various possible procedures. For other aspects of spectrum estimation, other time series procedures and relevant computational procedures the reader is referred to the short bibliography at the end of this paper.

First, we should provide some motivation for the use of frequency analysis. Any series of data points which may represent the values of some variable and which are equally spaced in time may be exactly described by a sum of periodic functions. This description in terms of periodic functions may be thought of as a description which depends on the set of frequencies. Thus, every set of values of a variable at equally spaced time points may be uniquely represented by a set of trigonometric functions with specified frequencies and amplitudes. Using the representation in terms of frequencies has several advantages:

1) Any tendency in the original data for periodic variations to occur in particular frequency ranges may be made more obvious in this representation.

2) Under appropriate conditions, the coefficients which form the representation will be independent of each other. This makes the representation much "simpler" in that we may consider the coefficients separately.

3) Under appropriate statistical assumptions this representation provides a means of forming statistically reasonable estimates of the parameters of the process which may be generating the observed data.

The formal statement of the above discussion is as follows: define the *finite Fourier series* representation of a set of values of a variable as:

$$h_n(\omega) = \sum_{t=1}^n \exp[-i\omega t] x_t, \quad -\pi \leq \omega \leq \pi, \quad (1)$$

where x_t is the value of the variable at time point t , for which we have values for $t = 1, 2, \dots, n$, ω is the frequency which may be defined at any point on the range $-\pi$ to π , and the complex exponential $\exp[-i\omega t]$ is the complex-valued periodic function which may also be written $\exp[-i\omega t] = \cos \omega t - i \sin \omega t$. $\exp[-i\omega t]$ is a periodic function of t with period $2\pi/\omega$ as can be seen by substitution and use of standard trigonometric identities. We can also define the inverse Fourier series which gives us the values of x_t from $h_n(\omega)$:

$$x_t = \frac{1}{n} \sum_{j=0}^{n-1} \exp[i\omega_j t] h_n(\omega_j), \quad t = 1, 2, \dots, n \quad (2)$$

where

$$\omega_j = j\pi/n.$$

We can see that these expressions give mutually reciprocal transformations by substituting eq. (1) to (2) to yield

$$x_t = \frac{1}{n} \sum_{j=0}^{n-1} \exp[i\omega_j t] (\exp[-i\omega_j] x_1 + \exp[-i\omega_j 2] x_2 + \dots + \exp[-i\omega_j t] x_t + \dots + \exp[i\omega_j n] x_n). \quad (3)$$

For each value of t only the term

$$\sum_{j=0}^{n-1} \exp[i\omega_j t] \exp[-i\omega_j t] x_t \quad (4)$$

will be non-zero. It will be

$$\sum_{j=0}^{n-1} \exp[i\omega_j t - i\omega_j t] x_t = (n)x_t. \quad (5)$$

While we may define $h_n(\omega)$ for any value of ω , it is completely determined — given n data points — by its values at the frequency points given by

$$\omega_j = j\pi/n, \quad j = 0, \dots, n. \quad (6)$$

This permits us to decompose a series into at most $n+1$ uniquely defined frequency components. The periodogram may be defined as:

$$p_n(\omega) = \frac{2}{n} |h_n(\omega)|^2. \quad (7)$$

From eq. (1) it would appear that we could evaluate the periodogram at any value of ω . However, from the standpoint of statistical interpretation it is very important to keep in mind that the periodogram is uniquely defined by its values at the $n+1$ frequency points given by eq. (6). In fact, a more usual definition of the periodogram is:

$$p_n(\omega_j) = \frac{2}{n} \left| \sum_{t=1}^n x_t \exp[-2\pi i j t/n] \right|^2, \quad j = 0, 1, \dots, n. \quad (8)$$

This definition makes the set of uniquely defining frequencies explicit. It must be emphasized that since these frequencies are spaced at intervals of width $1/n$ the resolution of periodograms (and spectra) is limited to at *very* best frequency intervals of width not less than $1/n$. The implications of this fact will be discussed in more detail below.

Current computational techniques imply that the most direct way to compute a sample spectrum is to compute the periodogram ordinates using a fast Fourier transform algorithm and then smooth the periodogram to obtain a consistent estimate of the spectrum.

In order to appreciate the statistical consequences of computing a spectrum estimate from a sample of data we must consider the following. A sample of data can be considered as if there were an infinitely long sample multiplied by the function:

$$\begin{aligned} w_t &= 1, & 1 \leq t \leq T, \\ w_t &= 0, & \text{otherwise,} \end{aligned} \quad (9)$$

where T is the length of the time average used. If we form a spectrum estimate then

$$f_s(\omega) = \int_0^\pi |L(\omega - \lambda)|^2 f_t(\lambda) d\lambda, \quad (10)$$

where $L(\lambda)$ is the Fourier transform of eq. (9), $f_t(\lambda) = |K_t(\lambda)|^2$ is the spectrum of the infinitely long series, and $f_s(\omega)$ is the sample spectrum. It is important to remember that we must form estimates of $f_t(\omega)$ and $f_s(\omega)$. The quality of these estimates depends on, among other things, $L(\lambda)$ and $K_t(\lambda)$. Using this framework we can see that there are ways to improve the quality of the estimates by effectively changing $L(\lambda)$. In particular, instead of using the rectangular data window given by eq. (10) one can use a window which tapers more gradually to zero at each end. This has the effect of producing an $L(\lambda)$ which is smoother and permits much less contamination of the estimate at one frequency by variation from another neighboring frequency.

Now, if we consider the simple case in which $f_t(\omega)$ is given by

$$f_t(\omega) = K + h\delta(\omega - \omega_0), \quad (11)$$

where K = constant, $\delta(\omega_0)$ = delta function at ω_0 , and h = scale parameter, we can derive the shape of $f_s(\omega)$ and then discuss the precision with which we can determine ω_0 .

Initially we may take $L(\lambda)$ to be the Fourier transform of $w(t)$. Thus

$$|L(\lambda)|^2 = \left(T \frac{\sin \lambda \frac{1}{2} T}{\frac{1}{2} \lambda T} \right)^2. \quad (12)$$

In order to form an estimate of $f_s(\omega)$ using eq. (12) we would have to compute the periodogram of the data and use the squares of the periodogram ordinates. It is well known that this is a risky thing to do in general but if $h \gg K$ (i.e., our process is known to contain a very large deterministic periodic variation at the unknown frequency ω_0) this procedure may be reasonable. It is important to consider this procedure because it results in the highest possible resolution of frequencies at the cost of undesirable statistical properties for the estimates. In this case $f_s(\omega)$ takes the form

$$f_s(\omega) = K + h \left(T \frac{\sin(\omega - \omega_0) \frac{1}{2} T}{(\omega - \omega_0) \frac{1}{2} T} \right)^2. \quad (13)$$

This results in a spectrum with the shape shown in fig. 1.

A variety of measures of the width of a peak such as this have been proposed. As reasonable as any, for our purposes, is the half power point given by the distance along the x-axis between the two points for which:

$$\left(\frac{\sin(\omega - \omega_0) \frac{1}{2} T}{(\omega - \omega_0) \frac{1}{2} T} \right)^2 = \frac{1}{2}. \quad (14)$$

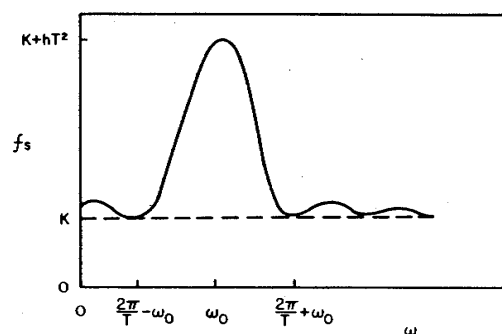


Fig. 1.

This distance is

$$2\Delta\omega \approx \frac{1}{T} \sin^{-1}\left(\frac{1}{2}\right), \quad (15)$$

where

$$\Delta\omega = |\omega - \omega_0|$$

or

$$2\Delta\omega \approx \frac{2}{3} \frac{\pi}{T}. \quad (16)$$

As an example, if we consider the data required so that $\Delta\omega = 1/720$ cycles/day (this is a shift of 2 min/cycle for a cycle with period one day), we find that

$$T_2 \approx 750 \text{ days}. \quad (17)$$

Given effectively error-free data and a very low variance at frequencies around the supposed delta function we could reasonably say that, given these data, the delta function appears to be located at

$$\omega_0 \pm \Delta\omega. \quad (18)$$

However, in this procedure we have assumed the existence of a delta function at some constant frequency. We should also consider the possibility that either the delta function changes its position or that the spectrum has a sharp but continuous peak. If we allow the possibility of some (random) variation in the (now not quite precisely) periodic term in the data then we must use up some of our data in order to average over this variability.

If we assume that the spectrum is continuous and locally smooth then, in the procedure which we have just outlined each estimate of the spectrum has a χ^2 distribution with two degrees of freedom. If we modify the procedure in order to in-

crease the degrees of freedom the data required for a given resolution increase proportionally with the degrees of freedom. In our example, if we want estimates with, say, eight degrees of freedom and still want to maintain $\Delta\omega = 2/1440$ then we must have

$$T_8 = 4T_2 = 3000 \text{ days}. \quad (19)$$

2.2. A frequency resolution nomogram

As the length or record required (T) is a linear function of the two variables, degrees of freedom (n) and resolution ($\Delta\omega$), we can present the relationships between these three variables using a tri-linear graph.

Fig. 2 presents the relationship

$$T = 750 \frac{n}{(\Delta\omega/\omega^2)}. \quad (20)$$

The axes have the following meanings:

Label	Variable	Range	Units
A-a	$\Delta\omega/\omega^2$	.5-128	min per cycle
B-b	n	2-1024	—
C-c	T	23-6000	days

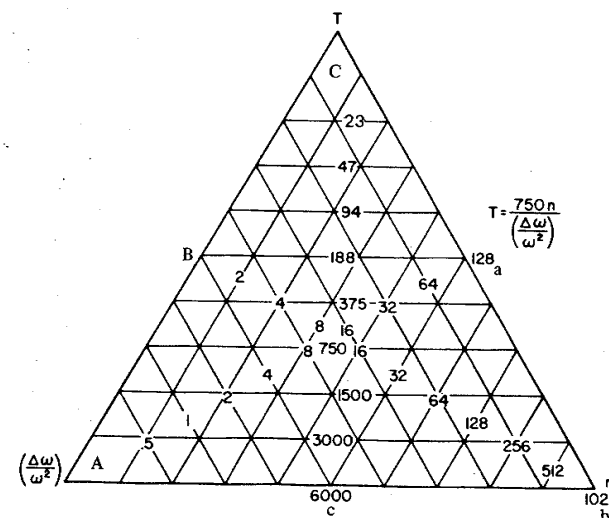


Fig. 2.

$\Delta\omega/\omega^2$ is derived from:

$$\Delta p = p_1 - p_2 = 2\pi \left(\frac{1}{\omega_1} - \frac{1}{\omega_2} \right) = \frac{2(\omega_2 - \omega_1)}{\omega_1 \omega_2} \approx \frac{2\pi \Delta\omega}{\omega^2},$$

$$(\omega_1 \omega_2 \gg \omega_2 - \omega_1),$$

and is used since it seems natural to measure resolution in time differences per cycle.

The intersections of the three perpendiculars to the three axes give the values of the three variables. Specification of the values of any two variables determines the value of the third.

2.3. Complex-demodulation

The estimation of spectra as described in the previous section requires forming averages over the entire length of the data. This procedure has an interpretation in terms of a population spectrum only if it is assumed that the process is wide-sense stationary. This may be a reasonable assumption for our data during periods when the environment is held constant. It obviously is not reasonable for the investigation of behavior at, and for some period after, a change in the environment. In order to investigate *changes* in behavior we have employed complex-demodulation.

The basic computational procedure is quite simple. We will state this procedure first, then indicate its theoretical motivation, and then show some practical properties.

A complex-demodulate of a time-series, x_t , at frequency ω , is defined by

$$x_t(\omega) = L(\exp[i\omega t] x_t), \quad (21)$$

where x_t is the observed time-series at $t = 1, \dots, n$, and $L(\)$ is an appropriate low-pass filter. If the process which generated x_t has a spectrum $f(\omega)$, then

$$f(\omega) = E(|x_t(\omega)|^2). \quad (22)$$

Therefore it seems natural to define a possible estimate of the sample spectrum by

$$f(\omega) = \frac{1}{n} \sum_{t=1}^n |x_t(\omega)|^2. \quad (23)$$

This equation clearly shows the relationship between a spectrum estimate and complex-demodulates. The average value of the amplitude of the complex-demodulate is a spectrum estimate at the frequency ω .

A complex demodulate at frequency ω may be thought of as a time series which contains the variation in the original time series which took place near the frequency ω . To see how this is true consider a time series:

$$x_t = A \cos \omega_0 t, \quad (24)$$

define $L(\omega)$ to be the Fourier transform of the function $L(\)$ which appears in eq. (21). For expositional simplicity we will take

$$L(\omega) = 1, \quad |\omega| \leq \Delta\omega, \quad (25)$$

$$L(\omega) = 0, \quad |\omega| > \Delta\omega.$$

Then, substituting eq. (24) into (21), we have

$$\begin{aligned} x_t(\omega) &= L\left\{\frac{1}{2} \exp[i\omega t] A (\exp[i\omega_0 t] + \exp[-i\omega_0 t])\right\} \\ &= L\left\{\frac{1}{2} A (\exp[i(\omega_0 + \omega)t] + \exp[i(\omega - \omega_0)t])\right\}. \end{aligned} \quad (26)$$

Now if $\omega - \omega_0 < \Delta\omega$ and $\omega_0 + \omega > \Delta\omega$ using eqs. (25) we have

$$x_t(\omega) = \frac{1}{2} A \exp[i(\omega - \omega_0)t]. \quad (27)$$

We can now rewrite eq. (27) in terms of amplitude and phase as

$$\begin{aligned} |x_t(\omega)|^2 &= \frac{1}{4} A^2, \\ \varphi_t(\omega) &= \tan^{-1} \frac{\text{Im}[x_t(\omega)]}{\text{Re}[x_t(\omega)]} = (\omega - \omega_0)t. \end{aligned} \quad (28)$$

3. Conclusions

In this section we will give a brief summary of our results and present figures which indicate some of the main findings.

Our results may be summarized as follows:

1) The computation of spectra and complex-demodulates yields results conforming to previous generalizations, that in nocturnal mammals the circadian period in LL is greater than the period in DD (see figs. 3, 4, 7, 10).

2) Previous evidence concerning the fact that there is a considerable period of adjustment when the photoperiod is changed is also supported by our computations. However, our evidence suggests that this adjustment period, and new equilibrium if any, are substantially more complex than had previously been indicated (see figs. 8, 9, 11).

3) The complex-demodulates indicate (see figs. 8, 9, 11) that there is substantial phase shift between the fundamental circadian frequency and its first few harmonics when a new photoperiod is initiated. In fact, in some cases one of the frequencies in the set of harmonics seems to lock onto the new photoperiod while others fail to do so. This appears to be a very interesting new finding, particularly in its relationship to the various coupled-oscillator hypotheses which have been suggested.

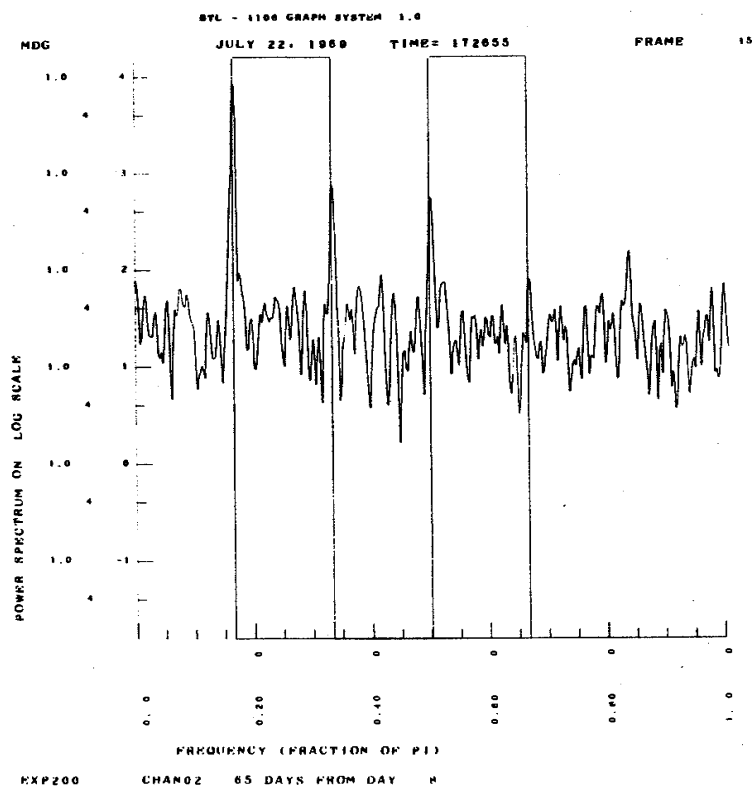


Fig. 3

(Vertical lines along horizontal axis in figs. 3-6 mark 24 hr period and harmonics.)

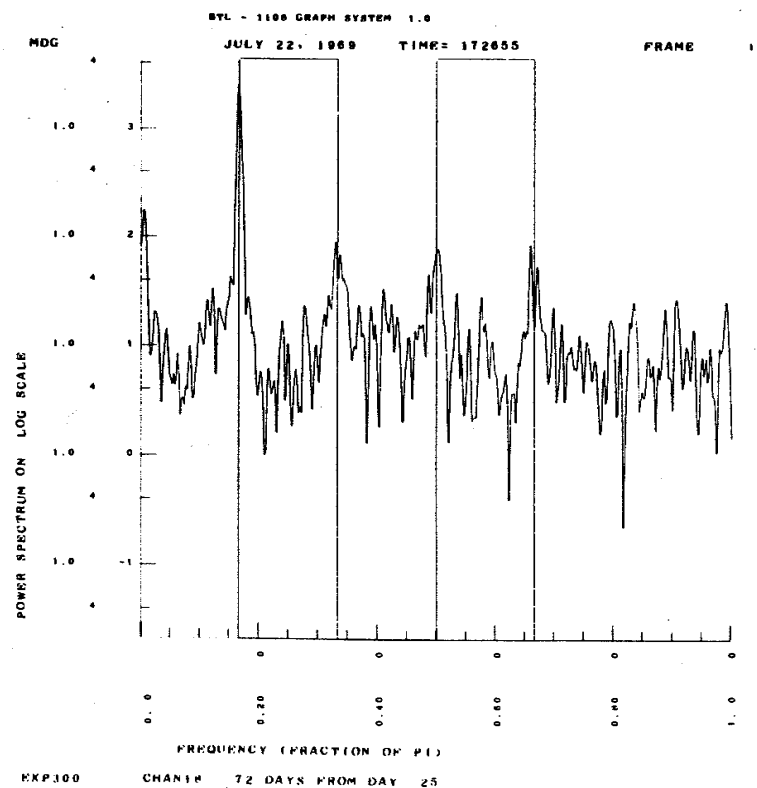


Fig. 4.

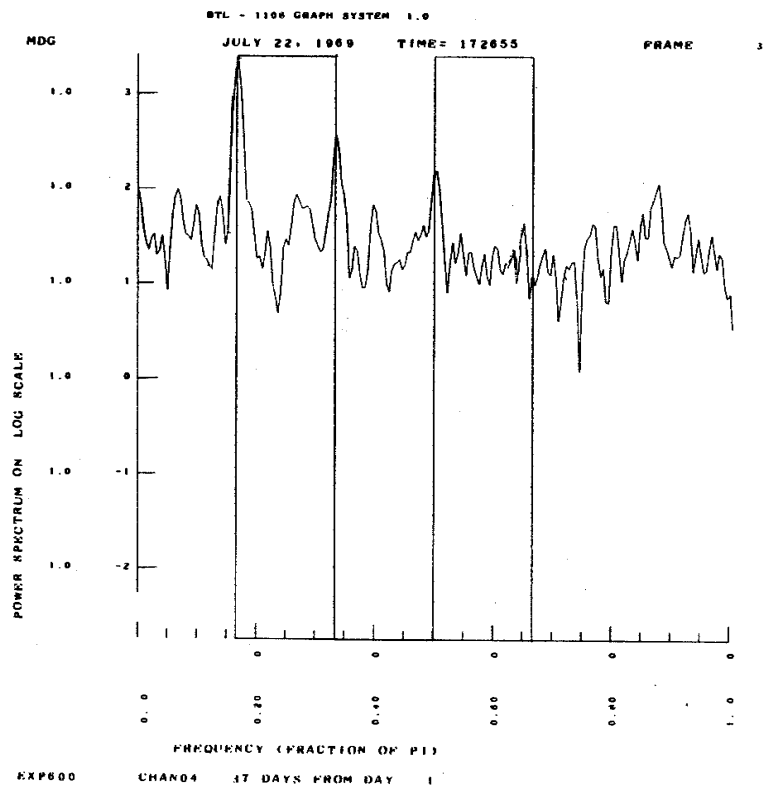


Fig. 5.

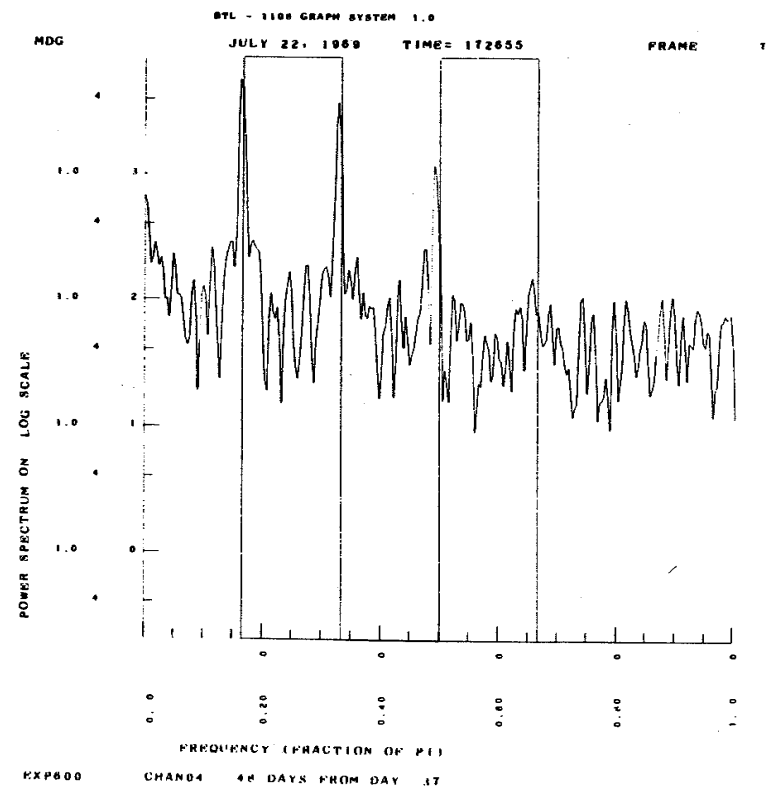


Fig. 6.

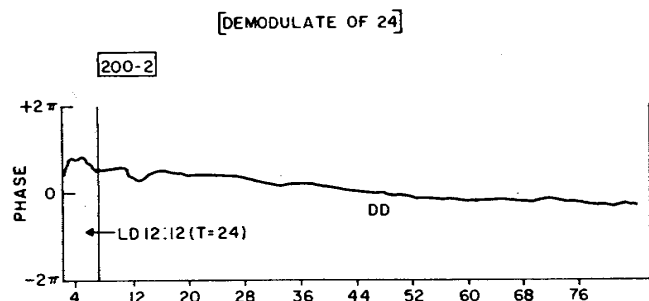


Fig. 7.

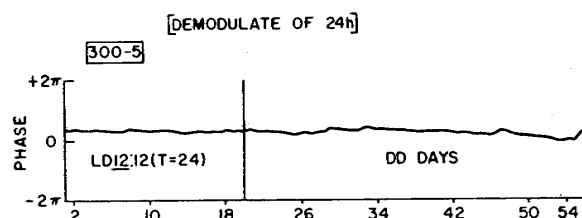


Fig. 8.

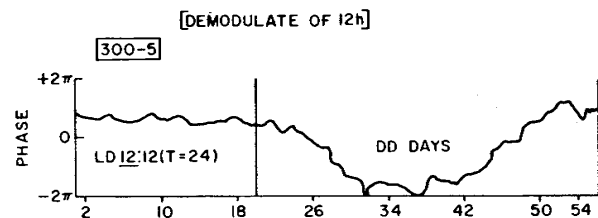


Fig. 9.

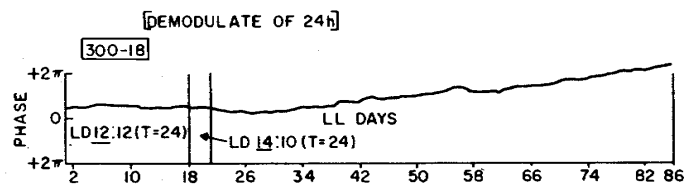


Fig. 10.

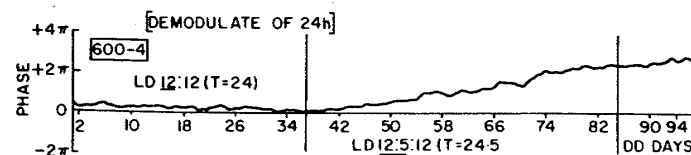


Fig. 11.

4) There appears to be a persistent overshoot in the adjustment to a new photoperiod if the new photoperiod is close to 24 hr. This overshoot is typically quite small, but it appears in many cases and over quite long time periods.

In conclusion we would like to remark that the body of data which we have been studying appears to be the most extensive, in terms of lengths of records and replication within experiments, and the most carefully controlled and instrumented that have yet been recorded. Our analyses have already provided new empirical evidence concerning hypotheses which attempt to explain the sources of circadian rhythms (particularly the coupled-oscillator hypotheses). We hope that as we and others continue to study these data substantial further insights will be obtained.

References

- [1] C.Bingham, M.D.Godfrey and J.W.Tukey, IEEE Transactions on Audio and Electroacoustics, Vol. AU-15, (1967) 56.
- [2] W.M.Gentleman and G.Sande, Fast Fourier Transforms - For Fun and Profit, Proc. Fall Joint Computer Conference (1966).
- [3] M.D.Godfrey, Appl. Stat. XIV (1965) 48.
- [4] M.D.Godfrey, Spectral Analysis of Time Series, ed. B.Harris (Wiley, New York, 1967) pp. 259-269.
- [5] M.B.Priestley, J. Sound. Vib. 6 (1) (1967) 86.
- [6] J.W.Tukey, in: Probability and Statistics (The Harald Cramér volume), ed. Ulf Grenander (Almqvist and Wiksell, Stockholm; Wiley, New York, 1959) pp. 300-330.

COSPAR

LIFE SCIENCES AND SPACE RESEARCH VIII

*Proceedings of the Open Meeting of Working Group V
at the Twelfth Plenary Meeting of COSPAR*

PRAGUE, 11-24 MAY, 1969

and

The Symposium on Biological Rhythms

PRAGUE, 13 MAY, 1969

and

The Symposium on Nutrition of Man in Space

PRAGUE, 14 MAY, 1969

Edited by

W. VISHNIAC and F. G. FAVORITE



1970

NORTH-HOLLAND PUBLISHING COMPANY - AMSTERDAM - LONDON

Part 4

Symposium on biological rhythms

59			
ditions		Stable circadian rhythms as a property of cell populations	
G.V.		E.E.Sel'kov, L.A.Sozinov	157
62		Circadian rhythm as response to the compulsory "constant" conditions	
otorula		W.B.Tshernyshev	168
68		Problems of latent desynchronosis	
ctivity		B.S.Alyakrinsky	171
		Evaluation of circadian dyschronism during transmeridian flights	
		A.Reinberg	172
77		Changes in the rhythm of some metabolic parameters under the influence	
arious		of sleep deprivation	
		E.Kuhn, V.Brodan, M.Brodanova, O.Schuck, J.Stribrna, J.Valek	175
82		Circadian rhythms and ionizing radiation effects	
od in		Yu.G.Grigoryev, N.G.Darenskaya, Yu.P.Druzhinin, S.S.Kuznetsov,	
		V.M.Seraya	176
90		The effects of electric fields on circadian rhythmicity in men	
nato-		R.Weaver	177
lies		Frequency spectra and cosinor for evaluating circadian rhythms in Rodent	
99		data and in man during Gemini and Vostok flights	
ctra-		F.Halberg	188
108		Data analysis of biological time series	
		M.D.Godfrey, H.Hoffman, R.Madden, C.S.Pittendrigh, S.Skopik	215
		Frequency measurement in time series data	
		E.Pöppel	234
		Circannual rhythmicity – evidence and theory	
		E.T.Pengelley, S.J.Asmundson	235
		Diurnal rhythm of the pituitary-adrenocortical response to stress: effect	
		of constant light and constant darkness	
119		J.Vernikos-Danellis, C.M.Winget, N.W.Hetherington	240
ated		Mechanisms of action of light on circadian rhythms in the monkey	
		C.M.Winget, L.S.Rosenblatt, C.W.DeRoshia, N.W.Hetherington	247

Part 5

Symposium on nutrition of man in space

142			
148		Problems of space nutrition of man	
		A.A.Pokrovsky, A.S.Ushakov	261
		Food technology problems related to space feeding	
		H.A.Hollender, M.V.Klicka, M.C.Smith	265